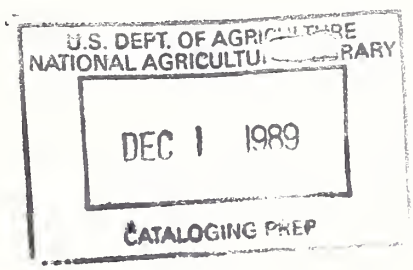
STATISTICAL ANALYSIS USING SPSS AT THE USDA-FORT COLLINS COMPUTER CENTER

BY
JAMES B. INGWERSEN

APPLICATION DOCUMENT SERIES
WSDG-AD-00002
DECEMBER 1980

USDA FOREST SERVICE
WATERSHED SYSTEMS DEVELOPMENT GROUP
3825 EAST MULBERRY STREET
FORT COLLINS, COLORADO 80524



List of Tables

STATISTICAL ANALYSIS USING SPSS AT THE USDA-FORT COLLINS COMPUTER CENTER

1.0 General Discussion

This document has been developed for individuals interested in conducting statistical analysis at the USDA-Fort Collins Computer Center (FCCC). Although there are numerous methods by which to accomplish this, the use of the SPSS (the Statistical Package for the Social Sciences) statistical package will be discussed here. It is assumed that you have access to FCCC documentation, an SPSS manual, and a terminal. You should recognize that the examples provided are not the only methods available, however, they will generate the desired results.

2.0 Building a Data File

To begin with, three ways to create a data file are illustrated. First, we will use the most common method of entering data on the system, the @FILE,UP statement. When using this approach, data should be punched on cards and placed in a deck with the following makeup.

```
@RUN,L RUN-ID,ACCOUNT #,QUALIFIER,S30,100
@FILE,UP QUALIFIER*YOURNAME.
place data cards here
@ENDF
@FIN
```

As far as the actual format of data on the cards goes, the following convention should be used. Identify records by site, date, time, etc., down the far left column just as you would number lines. Columns to the right of this should contain values for parameters in a fixed field format (that is the value of any given parameter always falls within a certain

range of columns). This makes checking of your data much easier than if it was in a free field format.

The second way to create a data file is with the use of the @DATA,I statement. This is most useful when data is stored on another medium before you create a system data file. For instance, if your terminal has a means of storing data, you can enter data into the memory and then read it into a system file at a later time. This process is highly useful when using word processors or the like. As far as the computer system is concerned, it is receiving "paper tape input." As a result, the @@PTI statement is used. The @@INQ statement provides a buffer for incoming data. A runstream would look something like the following:

```
@RUN RUN-ID,ACCOUNT #,QUALIFIER
@ASG,UP YOURFILE.
@DATA,I YOURFILE.
@@INQ
@@PTI
data transference here
enter a "control s" here
@@END
@END
@FIN
```

Since we are working with the terminal or word processor where the data is stored, these commands would be entered interactively (in demand mode).

The third way makes use of the editor. The @ED,I statement allows you to create a new data file. This statement is most effective in demand mode when there is not very much data to be entered. A runstream using this method would look like the following:

```
@RUN RUN-ID,ACCOUNT #,QUALIFIER
@ASG,UP YOURFILE.
@ED,I YOURFILE.
enter data here
enter carriage return
EXIT
@FIN
```

Some things should be kept in mind when using any of the above methods. The QUALIFIER of a file name will default to the PROJECT-ID that was entered in the RUN statement. The @FILE,UP statement, however, is an exception to this since the QUALIFIER must be explicitly stated. Also, it is important to keep your record length to a maximum of 80 columns. If a record (all data accompanying a site and date identifier) looks like it will exceed 80 columns, simply continue the record onto the next card. This format must be consistent throughout the data set, however, since a program or statistical package would look for two cards in every record.

2.1 Creating an SPSS Control Card Runstream.

At this point we should look at an actual SPSS control card runstream, specifically an example of a paired T-TEST. We will first look at the input data so the SPSS card INPUT FORMAT will make sense when we encounter it. Let's say that the example data in Table A is our input data and is located on the computer system in a file named QUALIFIER*EXAMPLE. The first thing to notice is the maximum number of columns that any variable occupies. In this case, specific conductance values occupy a maximum of three columns and fecal coliform values, two columns. Remember that according to our convention, this file has 100 lines with rank used as the identifier. The format of our data, therefore, would be F2.0,F3.0,F2.0 if the spaces between the data columns in the example were eliminated. If they were not, it would be F2.0,3X,F3.0,2X,F2.0.

Now that we are familiar with the input data, let's look at the SPSS run. The necessary statements to run a paired T-TEST are the following:

Table A. Populations of specific conductance (SC) and fecal coliform (FC), ranked in order of increasing magnitude, with means (μ) of 81.5 µmhos/cm and 23.6 counts/100 ml and standard deviations (σ) of 12.1 µmhos/cm and 12.5 counts/100 ml, respectively.

Rank	SC	FC	Rank	SC	FC	Rank	SC	FC	Rank	SC	FC	Rank	SC	FC
1	51	7	11	67	12	21	72	15	31	75	16	41	79	18
2	56	8	12	67	13	22	72	15	32	75	17	42	79	19
3	58	9	13	68	13	23	73	15	33	76	17	43	79	19
4	59	10	14	68	13	24	73	15	34	77	17	44	79	19
5	61	10	15	68	13	25	74	15	35	77	17	45	80	19
6	62	11	16	69	14	26	74	16	36	78	17	46	80	19
7	63	11	17	69	14	27	74	16	37	78	18	47	80	20
8	64	11	18	70	14	28	74	16	38	78	18	48	81	20
9	65	12	19	71	14	29	75	16	39	78	18	49	81	20
10	66	12	20	72	14	30	75	16	40	78	18	50	82	20

Rank	SC	FC	Rank	SC	FC	Rank	SC	FC	Rank	SC	FC	Rank	SC	FC
51	82	21	61	84	23	71	88	26	81	92	29	91	98	40
52	82	21	62	85	23	72	88	26	82	92	30	92	99	44
53	82	21	63	85	24	73	88	27	83	93	30	93	100	45
54	82	21	64	86	24	74	89	27	84	93	31	94	100	46
55	83	22	65	86	24	75	89	27	85	94	31	95	101	49
56	83	22	66	86	24	76	89	28	86	94	32	96	103	54
57	83	22	67	86	25	77	90	28	87	95	32	97	105	59
58	83	22	68	87	25	78	90	28	88	95	33	98	107	61
59	83	23	69	87	25	79	91	29	89	96	37	99	108	67
60	84	23	70	87	26	80	92	29	90	97	38	100	111	75

```
@ASG,A EXAMPLE.
@USE 8.,EXAMPLE.
@FC*GPL.SPSS
RUN NAME        T-TEST OF THE DATA IN EXAMPLE
INPUT MEDIUM    TAPE
N OF CASES      UNKNOWN
VARIABLE LIST   RANK,SC,FC
INPUT FORMAT    FIXED(F2.0,F3.0,F2.0)
T-TEST          PAIRS=SC WITH FC/
STATISTICS      ALL
READ INPUT DATA
FINISH
```

For a detailed explanation of the SPSS commands refer to an SPSS manual. Note that all SPSS program control identifiers are between columns 1 and 15 and instructions begin in column 16. One of the required cards in the series of SPSS control cards is the INPUT MEDIUM card. On this card you can inform SPSS whether your data is in a computer system file (TAPE) or on CARD. When you specify TAPE, SPSS will look on the input unit 8 for input data. Therefore, all SPSS runstreams using an online raw data set for input must have the following statements.

```
@ASG,A YOURNAME.
@USE 8.,YOURNAME.
```

This will enable your data file to act as input unit 8.

If you specify CARD on the INPUT MEDIUM card, SPSS will look for the data to be included within the runstream of SPSS control cards. Data cards would follow the SPSS control statement READ INPUT DATA in this case.

Runstreams similiar to the T-TEST example above can be seen in the Watershed System Development Group online userguide. To gain access to these examples, simply enter @XQT WSDU*WSDG.USERGUIDE. According to the FCCC naming convention everything to the left of the period will be interpreted as qualifier and filename and everything to the right as element name and version name. Please refer to FCCC documentation for explanation of file name conventions.

One way to create your own SPSS control card runstream is using the same @ED,I statement we used earlier. In this case, though, you can build an element. The reason we did not build an element for our data is because the @USE command does not apply to elements within a program file, only data files. The necessary statements to build an SPSS control card runstream using this method would be the following:

```
@ASG,UP SPSSRUN.
@ED,I SPSSRUN.WATER
enter carriage return
SET 16
MSCHAR X
enter carriage return
XASG,A DATAFILE.
XUSE 8.,DATAFILE.
XFC*GPL.SPSS
RUN NAME;MY SPSS ANALYSIS
enter remaining SPSS control card runstream here
enter carriage return
EXIT
```

The SET 16 statement sets a tab at column 16. Remember that all instructions applying to SPSS control cards must begin in or after column 16. Tabbing is accomplished by entering a semicolon (;). Note that this is used in the RUN NAME statement above. The MSCHAR statement sets a character which is to be translated to a master space (more commonly known as the "at" symbol, @) when it is input in column one of input lines in input mode. In this case, the master space character is an X. The reason this statement is needed is because a master space entered in column one while in the editor will be interpreted as an EXIT.

Another effective way to build an element is with the use of the @ELT,ID statement. The commands issued using this method would be the following:

```
@ASG,UP SPSSRUN.
@ELT,ID SPSSRUN.WATER
@ASG,A DATAFILE.
@USE 8.,DATAFILE.
@FC*GPL.SPSS
RUN NAME        MY SPSS ANALYSIS
INPUT MEDIUM    TAPE...
enter remaining SPSS control card runstream here
@END
```

Both the @ED,I and the @ELT,ID create elements. You should realize at this point why data is placed in a data file. For convention, our runstream is placed in an element of a program file.

3.0 Executing an SPSS Job.

Since your runstream would probably be created interactively, the easiest way to execute your SPSS job is by adding your SPSS control card runstream to your current interactive run. This is accomplished by using the @ADD statement. For instance, if your runstream is in an element named SPSSRUN.WATER, you would simply enter @ADD SPSSRUN.WATER during your demand run. The output of the program will then be seen at your terminal as the system processes the commands in the runstream. If it appears to be successful, depress the break key. This will stop the output temporarily and the system will respond OUTPUT INTERRUPT. Enter an @@X TIO, which will terminate execution of the SPSS program. The system will continue to print out a couple more lines since these are in an intermediate buffer.

To send the output to a high speed printer, enter the following commands.

```
@ASG,UP PFILE.
@BRKPT PRINT$,PFILE
@ADD SPSSRUN.WATER
@BRKPT PRINT$
@FREE PFILE.
@SYM PFILE,# of copies, site id
```

If there were any errors in the run as denoted by SPSS error messages or system messages, make necessary corrections to your SPSS control card element with the editor. To edit SPSSRUN.WATER for instance, enter @ED,U SPSSRUN.WATER. The U signifies update. Once you are sure that the errors have been corrected and checked, enter the statement @ADD SPSSRUN.WATER again.

As mentioned earlier, the primary mechanism for Watershed Systems Development Group support at the Fort Collins Computer Center is through the use of the program named USERGUIDE. Upon execution of the USERGUIDE program, one can obtain an index of the different types of documentation available from the WSDG. One of these types is the userguide series. These examples are online and can be obtained either directly over a low speed terminal or printed at a high speed printer.

The USERGUIDE program is self-explanatory when being executed and is simple to use. It is highly encouraged that you occasionally review the various lists of current support documentation indexed by the USERGUIDE program. SPSS documentation is unique to other documentation in USERGUIDE in that it is divided into separate subjects corresponding to the procedures and use involved with different SPSS subprograms. General information about SPSS control cards and program execution information which applies to all SPSS subprograms can also be obtained in USERGUIDE.

To execute USERGUIDE, enter the following command:

```
@XQT WSDU*WSDG.USERGUIDE
```

A brief summary of SPSS control cards and program explanations is included in Appendix A.

APPENDIX A

SPSS QUICK REFERENCE GUIDE

```
1                    16
ASSIGN MISSING       varlist1(missing values)/varlist2 . . . / . . .
```

Function (p. 115) Assigns a missing value for computed variables. When a case contains missing values for any one of the variables used to calculate a computed variable, that case will assume the assigned missing value for the computed variable.

Status Optional.

```
1                    16
COMMENT              any text
```

Function (p. 75) Allows insertion of user comments anywhere in the control card deck except between a control card and its continuation or within the data deck.

Status Optional.

```
1                    16
[*] COMPUTE          computed variable=arithmetic expression
```

Function (p. 96; pp. 160, 161†) Creates new variables or new values for existing variables by combining existing variables according to one or more of the functions listed below. Only one transformation may be requested per COMPUTE card.

Arithmetic operators

+	addition	/	division
**	exponentiation	*	multiplication
–	subtraction		

Special Functions

ABS	Absolute value	ATAN	Arctangent
COS	Cosine	EXP	Exponential (e^{arg})
LG10	Base 10 logarithm	LN	Natural logarithm
RND	Round to whole number	SIN	Sine
SQRT	Square root	TRUNC	Truncated value (whole number without rounding)
MOD10	Remainder of division by 10		

NORMAL (argument) Generates a normally distributed variable; mean = 0. The (argument) specifies the standard deviation.

UNIFORM (argument) Generates a uniformly distributed variable; lower limit = 0. The (argument) specifies the upper limit.

YRMODA (yr,mo,day) Converts dates to day numbers for computing time intervals. Years 0–99 are 20th century; otherwise, years range 1582–47516. Months range 1–13; days range 0–31.

Limitations

- Variables SEQNUM, SUBFILE, and CASWGT may not be given new values.
- A maximum of 250 elements (variable names, keywords, relational operators, arithmetic operators, numeric constants, left or right parentheses, equal sign) may appear on the control card.

Status Optional.

Fixed-Column Format

```
1                    16
DATA LIST            FIXED {(1)                    } /recnumber varlist1
                           {(records per case)     }
                     start position[–end position {(0)} ]
                                                  {(n)}
                                                  {(A)}
                     varlist2 . . . /recnumber . . .
```

Binary Format

```
1                    16
DATA LIST            BINARY {(1)                   } /recnumber varlist1
                            {(records per case)    }
                     start position[–end position]
                     varlist2   ./recnumber
```

Function (p. 49) Names variables and specifies their formats. Position specifications may note the location of one variable or the range within which a series of variables of the same column width and type occur. FIXED assumes that all variables are numeric with no implied decimals, unless otherwise informed by variable type notations "A" (alphanumeric) or "n" ("n" decimal places). BINARY assumes that all variables are 4 byte internal format floating point numbers.

Status Alternative to the combination VARIABLE LIST and INPUT FORMAT for entering raw data.

```
1                    16
END INPUT DATA
```

Function (p. 174†) Signals end of raw data embedded in the SPSS control deck. It is placed at the end of the raw data (and before the next SPSS control card).

Status Conditional. Required with N OF CASES UNKNOWN and INPUT MEDIUM CARD.

```
1                    16
FILE NAME            'filename [file label]
```

Function (p. 36) Names and labels a file. The filename is a standard SPSS 8-character name; the optional label may be up to 40 characters long.

Status Optional.

```
1                    16
FINISH
```

Function (p. 73; p. 216†) Terminates processing.

Status Optional

```
1                    16
GET FILE             'filename
```

Function (p. 85) Retrieves a system file.

Status Conditional. Required whenever an existing system file is accessed.

1
[*] IF

16
(logical expression) computed variable=
arithmetic expression

Function (p. 101; p. 173†) Creates new variables using COMPUTE functions (listed on page 3 of this manual) provided the logical expression is true (logical operators are listed below). Only one transformation may be entered per IF card.

GE	Greater than or equal to	NE	Not equal to
LE	Less than or equal to	AND	
GT	Greater than	OR	
EQ (or =)	Equal to	NOT	

Limitations

- Variables SEQNUM, SUBFILE, and CASWGT may not be given new values.
- A maximum of 250 elements (variable names, keywords, logical operators, relational operators, arithmetic operators, numeric constants, left or right parentheses, equal sign), may appear on a card.

Status Optional.

1
INPUT FORMAT

16
{FIXED / BINARY} (format list)
or
FREEFIELD

Function (p. 41; pp. 176, 177†) Describes data formats: organization of cases, type and location of variables. No format is specified for FREEFIELD input.

Fixed-column Format Elements

nFw.d	Numeric item of width "w" columns counting sign and decimal if punched and "d" digits interpreted or punched right of the implied decimal point.
nAw	Alphanumeric variable of column width "w". Maximum "w" = 4.
nBw.d	Binary format item of width "w" in bytes (either 2 or 4) and "d" decimal digits to the right of the implied decimal point.
nCw.d	Commercial or packed decimal format item of width "w" in bytes and "d" digits to the right of the implied decimal point.
nRw	Real or internal format floating point item of width "w" in bytes.
nZw.d	Zoned decimal format item of width "w" in bytes and "d" decimal places to the right of the implied decimal point.
nX	Skip "n" columns.
Tn	Transfer to column "n".
/	Skip to next record.
n ( )	Repeat the operation in parentheses "n" times.

Binary Format Elements

nV	"n" adjacent variables on the file to be input. A "V" must be entered for every variable named or implied.
nS	Skip "n" adjacent variables.
Pn	Proceed to the "nth" variable.
/	Skip to the next record.
n ( )	Repeat the operation in parentheses "n" times.

Status Conditional. Required on all runs containing VARIABLE LIST, ADD VARIABLES, ADD SUBFILES, or ADD CASES.

1
INPUT MEDIUM

16
CARD
or
{DISK / TAPE / OTHER}

Function (p. 39; p. 175†) Identifies the medium through which data are to be entered. Enter keyword CARD when submitting data on cards within the SPSS control deck or from terminal. Keywords other than CARD are essentially synonymous and require an operating system control card defining the raw input file. The default medium is CARD.

Status Conditional. Required on all runs reading raw data.

1
LIST CASES

16
CASES={10 / n}/VARIABLES={ALL / varlist}

Function (p. 137) Lists values of the selected variables for the first "n" cases. If the CASES=n parameter is omitted, values for the first 10 cases of each subfile, or for the first 10 cases in the file if there is no subfile processing, are printed. Must directly precede a procedure card which activates processing of the file.

Status Optional.

1
MISSING VALUES

16
varlist$_1$ (value list$_1$) varlist$_2$ . . .

Function (p. 57; p. 220†) Defines missing values. Keyword THRU may be used to indicate a range of missing values; keywords LOWEST (LO) and HIGHEST (HI) may be used in place of values with THRU.

Limitations

- A maximum of three missing specifications may be entered per list. This includes the upper or lower ends of a range as well as single values.
- A maximum of 250 elements (names, values and special delimiters) may appear on a MISSING VALUES card and its continuations.

Status Optional.

1
N OF CASES

16
{UNKNOWN / n of cases}

Function (p. 40; pp. 174, 176†) Specifies the number of cases in the file.

Status Optional.

1
READ INPUT DATA 16

Function (p. 67; p. 175†) Instructs the system to begin reading input data.

Status Optional unless INPUT MEDIUM is CARD.

1
[*] **RECODE** 16
varlist$_1$ ({value list$_1$ / BLANK} =newvalue$_1$)
or
(CONVERT)
({value list$_2$ / ELSE / BLANK} =newvalue$_2$) ../varlist$_2$
or
(CONVERT)

Function (p. 90) Recodes variable values. Adjacent values may be listed implicitly with keyword THRU. Keywords LOWEST (LO) and HIGHEST (HI) may be used instead of numeric values. ELSE recodes all values not otherwise assigned. BLANK assigns values to blank data which otherwise are treated as zero. CONVERT recodes alphanumeric variables into numeric variables.

Limitations

- Variables SEQNUM, SUBFILE, and CASWGT may not be recoded.
- No more than 250 elements (variable names, keywords, individual values, left or right parentheses, and equal sign) may be included on a card. Names implied by the TO convention are not counted.
- A maximum of 400 variables may be named or implied on the entire set of RECODE and COUNT cards in effect for a procedure (800 for Version M).

Status Optional.

1
RUN NAME 16
run label

Function (p. 72) Prints a run name of up to 64 characters in length at the top of each page of output.

Status Optional.

1
SAVE FILE 16
[filename[file label]]

Function (p. 81; p. 223†) Saves raw input file as a system file; the optional specification field allows renaming and relabeling of the file.

Status Required whenever a file generated on a run or modifications to an existing system file are to be permanently retained.

1
[*] **SELECT IF** 16
(logical expression)

Function (p. 128; p. 160†) Selects for processing only those cases which meet the criteria specified in the logical expression (see page 7 of this guide for a list of the logical operators). Cases must satisfy the criteria on each SELECT IF card included in a run in order to be selected.

Status Optional.

1
STATISTICS 16
{statistics number list / ALL}

Function (p. 66) Requests procedure statistics.

Status Optional.

1
VALUE LABELS 16
{varlist$_1$ / ALL} (value$_1$)label$_1$ (value$_2$)label$_2$ ../
varlist$_2$ ../...

Function (p. 59) Associates a label of up to 20 characters with each value for the variables specified.

Status Optional.

1
VAR LABELS 16
varname$_1$ label$_1$/varname$_2$ label$_2$ ...

Function (p. 62) Associates a variable label of up to 40 characters with the variables specified.

Status Optional.

1
VARIABLE LIST 16
varlist

Function (p. 36) Names variables in the order defined on the subsequent INPUT FORMAT card. Up to 8 characters, the first of which must be alphabetic, are allowed per name. Names may be listed with the TO convention when they are of the form "alphannn TO alphammm" where mmm is greater than nnn

Status Either VARIABLE LIST or DATA LIST is required when the data are to be read from a raw data file.

```
1                    16
ANOVA                dependent varlist1 BY independent varlist1(min,max)
                     independent varlist2(min,max)    [WITH covariate list] /
                     dependent varlist2 .../...
```

Reference (p. 410; p. 206†) Note: The default method is classic.

Options (p. 418; p. 206†)

1 Includes missing data.
2 Suppresses value labels.
3 Deletes all interaction terms from the model.
4 Deletes three-way and higher interaction terms.
5 Deletes four-way and higher interaction terms.
6 Deletes five-way interaction terms.
7 Processes covariates concurrently with main effects for nonmetric factors.
8 Processes covariates after main effects for nonmetric factors.
9 Regression approach. Overrides Options 7 and 8.
10 Hierarchical approach.
11 Prints with an 80 character width.

Statistics (p. 419; p. 207†)

1 Multiple classification analysis. Will not be produced when Option 9 is specified.
2 Unstandardized partial regression coefficients for the covariates.
3 Prints the means and the counts table. Not available with Option 9.

Limitations (p. 419)

- A maximum of 5 dependent variables, 5 independent variables, and 5 covariates.
- A maximum of 5 analysis lists (sets of independent and dependent variables, with covariates).
- The combined number of categories for all nonmetric factors of an ANOVA analysis list plus the number of covariates must be less than the sample size.

```
Integer Mode
1                    16
BREAKDOWN            VARIABLES=varlist1 {(min,max)          } varlist2 .../
                                        {(LOWEST,HIGHEST)}

                     {TABLES     } =varlist BY varlist BY .../... varlist...
                     {CROSSBREAK}
General Mode
1                    16
BREAKDOWN            TABLES=varlist BY varlist BY .../varlist BY .../...
```

Reference (p. 249; p. 214†)

Options (p. 257)

1 Includes missing data.
2 Excludes missing data from the dependent variable only.
3 Suppresses printing of labels.
4 Prints BREAKDOWN tables in modified tree-diagram format. Available in general mode only.

Options 5–8 apply to CROSSBREAK= only (p. 266)

5 Suppresses printing of cell frequencies.
6 Suppresses printing of sums.
7 Suppresses printing of standard deviations.
8 Suppresses value labels, but prints variable labels.

Statistics (p. 258)

Statistics 1 and 2 do not apply to CROSSBREAK=

1 One-way analysis of variance.
2 Calculates a test of linearity. Statistic 1 must also be requested.

Statistics 3–12 apply to CROSSBREAK= only (p. 266)

3 Chi-square
4 Phi for 2 x 2 tables, Cramer's V for larger tables
5 Contingency coefficient
6 Lambda, symmetric and asymmetric
7 Uncertainty coefficient, symmetric and asymmetric
8 Kendall's tau b
9 Kendall's tau c
10 Conditional gamma
11 Somers's D, symmetric and asymmetric
12 Eta

Limitations: Integer Mode (p. 262; p. 214†)

- A maximum of 100 variables may be named or implied on the VARIABLES= list.
- A maximum of 100 variables may be named or implied on the TABLES= list.
- A maximum of 30 TABLES= lists may be entered.
- A maximum of 100 tables may be requested.
- A BREAKDOWN table may have a maximum of 6 dimensions: dependent variable plus 5 independent variables.

Limitations: General Mode (p. 261)

- A maximum of 200 variable names may be entered or implied.
- A maximum of 30 TABLES= lists may be entered.
- A maximum of 250 tables may be requested.
- Value labels in excess of 200 per BREAKDOWN table are not printed.
- A BREAKDOWN table may have a maximum of 6 dimensions: dependent variable plus 5 independent variables.

```
1                16
CONDESCRIPTIVE  {varlist}
                {ALL   }
```

Reference (p. 185; p. 215†)

Options (p. 189)

1 Includes missing data.

2 Suppresses printing of variable labels.

3 Writes out Z scores.

4 Prints a reference index.

Statistics (p. 189)

1 Mean

2 Standard error

5 Standard deviation

6 Variance

7 Kurtosis

8 Skewness

9 Range

10 Minimum

11 Maximum

12 Sum

Limitations (p. 190)

- A maximum of 250 variable names may be entered. Variable names implied by the TO convention do not count toward this total.
- A maximum of 500 variables may be named or implied.

```
Integer Mode
1                16
CROSSTABS        VARIABLES=varlist(min,max) varlist(min,max)   /
                 TABLES=varlist BY varlist   /varlist BY varlist   .../

General Mode
1                16
CROSSTABS        TABLES=varlist BY varlist   /varlist BY varlist   /
```

Reference (p. 230; p. 200 †)

Options (p. 241; p. 200†)

1 Includes missing data.

2 Suppresses printing of labels.

3 Suppresses printing of row percents.

4 Suppresses printing of column percents.

5 Suppresses printing of total percents.

6 Suppresses value labels, but prints variable labels. Not available in general mode.

7 Includes missing values in tables but not in calculation of statistics or of row and column percents. Not available in general mode.

8 Prints tables ordered on row variable values, highest to lowest. Not available in general mode.

9 Prints an index of tables.

10 Output non-empty cell contents.

11 Output all cell contents.

12 Supress printed output.

Statistics 9p. 242; p. 205†)

1 Chi-square

2 Phi for 2 x 2 tables, Cramer's V for larger tables

3 Contingency coefficient

4 Lambda, symmetric and asymmetric

5 Uncertainty coefficient, symmetric and asymmetric

6 Kendall's tau b

7 Kendall's tau c

8 Gamma (partial and zero-order gamma available in integer mode only)

9 Somers's D, symmetric and asymmetric

10 Eta (available for numeric data only)

11 Pearson's r; available for numeric data only

Limitations: Integer Mode (p. 244)

- A maximum of 100 variables may be named or implied on the VARIABLES= list.
- A maximum of 100 variables may be named or implied on the TABLES= list.
- A maximum of 20 TABLES= lists may be entered.
- A crosstabulation table may have a maximum of 8 dimensions: 6 control variables plus the independent and dependent variables.

Limitations: General Mode (p. 243)

- A maximum of 200 variables may be named or implied on all TABLES= lists combined.
- A maximum of 250 values will be printed for each variable.
- A maximum of 20 TABLES= lists may be entered.
- A crosstabulation table may have a maximum of 10 dimensions: 8 control variables plus independent and dependent variables.

```
1
DISCRIMINANT    18
  GROUPS= {varname(min,max)} /VARIABLES=varlist/
          {SUBFILES       }
  SELECT=varname(value)/
  ANALYSIS=varlist1 {(1)*  } [,varlist2 .../ANALYSIS=.../.../]
                    {(level)}
  METHOD= DIRECT    /TOLERANCE= {.001}/
          WILKS                 {t   }
          MAHAL
          MAXMINF
          MINRESID
          RAO*
  MAXSTEPS= {2v}/ FIN= {1.0}/FOUT= {1.0}/PIN= {1.0}/
            {m }       {fi }       {fo }      {pi }
  POUT= {1.0}/VIN= {0 }/
        {po  }     {vi}
  FUNCTIONS= {g-1,100.0,1.0}/PRIORS= {EQUAL    }/
             {nf  cp   sig }         {SIZE     }
                                     {value list}
  [ANALYSIS= ...]
```

Reference (p. 448; p. 184†)

Note: Only GROUPS= and VARIABLES= must be coded; brackets indicating that all other keywords are optional have been removed for readability.

Options (p. 456; p. 189†)

1 Includes missing data.
2. Includes cases with missing data during classification.
3 Suppresses printing of step-by-step output.
4 Suppresses printing of the step-by-step summary table.
5 Prints the classification results table.
6 Prints discriminant scores and classification information for each case.
7 Prints a single scatterplot of cases.
8 Prints a separate scatterplot for each group.
9 Suppresses classification phase features for cases initially unclassified.
10 Prints a territorial map.
11 Prints unstandardized discriminant function coefficients.
12 Prints classification function coefficients.
13 VARIMAX rotation of discriminant functions.
14 Individual group covariance matrices are used for classification rather than the pooled within-groups covariance matrix.
15 Writes out matrix materials.
16 Signals matrix input.
17 Writes discriminant scores on the raw output file.
18 Writes membership probabilities for all groups.
19 Writes actual group and classified group numbers.

Statistics (p. 459; p. 197†)

1 Means
2 Standard deviations
3 Pooled within-groups covariance matrix
4 Pooled within-groups correlation matrix
5 Matrix of pairwise F ratios (available only with stepwise method)
6 Univariate F ratios
7 Test for equality of group covariance matrices (Box's M)
8 Group covariance matrices
9 Prints a total covariance matrix.
10 Prints the structure matrix.

Limitations (p. 461; p. 197†)

- Pairwise deletion of missing data is not available
- Only one GROUPS=specification, one SELECT= specification, and one VARIABLES=specification may be entered.

```
1
FACTOR    18
  VARIABLES=varlist/TYPE= {PA2     }/
                          {PA1     }
                          {RAO     }
                          {ALPHA   }
                          {IMAGE   }
                          {BYPASS  }
  DIAGONAL= {1.0       } /[NFACTORS=n]/
            {value list}
  MINEIGEN= {1.0}/ITERATE= {25}/STOPFACT= {.001}/
            {n  }          {n }           {n   }
  ROTATE= {VARIMAX  } /DELTA= {0}/
          {QUARTIMAX}         {n}
          {EQUIMAX  }
          {OBLIQUE  }
          {NOROTATE }
  [FACSCORE[=mdrp]]/VARIABLES=
```

Reference (p. 490; p. 216†)

Options (p. 503)

1 Includes missing data.
2 Excludes missing data pairwise.
3 Input of correlation matrix.
4 Input of factor matrix and communalities.
5 Writes out correlation matrix of all variables used in the factoring.
6 Writes out factor matrix and associated communalities.
7 Writes out the factor-score coefficient matrix.
8 Writes out means and standard deviations of the variables on the first VARIABLES= list.
9 Indexes the number and order of the variables on the input correlation matrix based on the VARIABLE LIST card.
10 Creates factor scores which are a weighted product of the existing data, if missing data replacement has been requested.
11 Records sequencing information in columns 1–20 of each factor-score record.

Statistics (p. 506)

1 Valid n's, means, and standard deviations
2 Correlation matrix
3 Inverse of the correlation matrix and its determinant
4 Eigenvalues, proportion of common variance, initial communality estimates, and final communalities
5 Initial unrotated factor matrix
6 Rotated factor matrix and transformation matrix
7 Factor-score coefficient matrix (not available when Option 4 is in effect)
8 Plot of variables on factor axes (not available for oblique rotation)

Limitations (p. 507)

- A maximum of 100 variables may be entered on a VARIABLES= list.
- All variables specified on second and subsequent VARIABLES= list must also be specified on the first VARIABLES= list.
- When subfile groups are being processed separately, only one VARIABLES= list may be entered.
- Factor scores for each case cannot be output when matrix input is used.

Integer Mode

```
1                16
FREQUENCIES      INTEGER=(varlist1 | ALL) (min,max)/
                 varlist2 (min,max) .../
```

General Mode

```
1                16
FREQUENCIES      GENERAL=(varlist | ALL)
```

Reference (p. 194; p. 216†)

Options (p. 200; p. 217†)

1 Includes missing data.
2 Suppresses value labels.
3 Prints output in an 8½" x 11" space.
4 Writes out all output.
5 Prints all requested frequency tables in condensed format.
6 Prints all requested frequency tables which will not fit on one page of output in condensed format.
7 Suppresses printing of frequency tables.
8 Prints a histogram for each variable.
9 Prints an index of tables.
10 Prints tables in descending order of values.*
11 Prints tables in descending order of frequency.*
12 Prints tables in ascending order of frequency.*

* Available in general mode only.

Statistics (p. 201)

1 Mean
2 Standard error
3 Median
4 Mode
5 Standard deviation
6 Variance
7 Kurtosis
8 Skewness
9 Range
10 Minimum
11 Maximum

Limitations (p. 201; p. 217†)

- A maximum of 500 variables may be named or implied.
- A maximum of 32767 values over all variables may be referenced.

```
1                16
NPAR TESTS       CHI-SQUARE=varlist(min,max)/
                 [EXPECTED=(EQUAL | proportions)]/
                 K-S (UNIFORM(min,max) | NORMAL(mean,s.d.) | POISSON(mean))=varlist/
                 RUNS((MEAN | MEDIAN | MODE | value))=varlist/
                 (MCNEMAR | SIGN | WILCOXON)=varlist WITH varlist/
                 (COCHRAN | FRIEDMAN)=varlist/
                 (MEDIAN(value) | M-W | K-S | W-W | MOSES(value) | K-W)=dependent varlist BY
                 independent var(value1,value2)/
```

Reference (p. 66†)

Options (p. 96†)

1 Includes missing data.
2 Excludes missing data listwise.
3 Sequential pairing of variable list. (See p. 17†)
4 Random sample of cases is selected if not enough space is available to process all cases.
5 Printed output is restricted to 75 columns.

Statistics (p. 97†)

1 Mean, maximum, minimum, standard deviation, and number of cases.

Limitations (p. 97†)

- Maximum of 200 test specifications; for tests CHI-SQUARE=, K-S= (one-sample), and RUNS()=, each variable tested counts as one; for all other tests, each test is counted as one, regardless of the number of variables named.

```
1                16
ONEWAY           dependent varlist BY independent var(min,max)/
                 [POLYNOMIAL=n/] [CONTRAST=coefficient list/
                 CONTRAST= .../.../] [RANGES={LSD, DUNCAN, SNK, TUKEYB, TUKEY, LSDMOO, SCHEFFE, or ranges values}{.05, alpha}]/
```

Reference (p. 422; p. 209†)

Options (p. 429; p. 209†)

1 Includes missing data.

2 Excludes missing data listwise.

3 Suppresses printing of variable labels.

4 Writes out the number of cases, mean, and standard deviation for each group.

6 Uses the first 8 characters of the independent variable value labels as group labels.

7 Input of category frequencies, means, and standard deviations instead of raw data.

8 Input of category frequencies, means, pooled variance, and degrees of freedom for the pooled variance.

10 Use the harmonic mean for all groups in range tests.

Statistics (p. 430; p. 210†)

1 Number of cases for each category, means, standard deviations, standard errors, minimum, maximum, and 95% confidence interval for the mean

2 Fixed- and random-effects measures

3 Homogeneity of variance statistics: Cochran's C, Bartlett-Box F, and Hartley F-max

Limitations (p. 430; p. 211†)

- A maximum of 20 dependent variables may appear.
- Only one independent variable may appear on the ONEWAY card; it must be integer-valued and may have a maximum of 50 categories included in the analysis.
- Only one POLYNOMIAL= specification is allowed. Up to 10 CONTRAST= specifications and 10 RANGES= specifications are allowed.
- Not all alpha values may be used with any RANGES= test.

```
1                16
REGRESSION       VARIABLES=varlist/REGRESSION=dependent var
                 {(80,.01,.001)}{(n, F, T)} WITH independent varlist
                 (inclusion level)  independent varlist {(1)}{(inclusion level)}
                 [RESID={mdrp}{0}]/REGRESSION= /VARIABLES=
```

Reference (p. 342; p. 223†)

Options (p. 352)

1 Includes missing data.

2 Excludes missing data pairwise.

3 Suppresses printing of variable labels.

4 Signals matrix input.

5 Input of means and standard deviations before input of the correlation matrix.

6 Suppresses printing of step-by-step output.

7 Suppresses printing of the summary table.

8 Writes out correlation matrices.

9 Indexes the correlation matrix by VARIABLE LIST card.

10 Enters sequencing information on residuals in columns 1 through 20 of each record in the raw output file. Residuals and predicted Y values are written in 6F10.6 format.

11 Writes out standardized residuals.

12 Writes out standardized predicted Y values.

13 Weights existing data to create standardized predictors. May be used only when Option 2 and missing data replacement have been requested (mdrp greater than zero).

14 Suppresses printing of axes on the plots of standardized predictor versus standardized residual obtained with Statistic 6.

15 Writes out means and standard deviations. Cannot be used with matrix.

Statistics (p. 355)

1 Prints a correlation matrix for each VARIABLES= list. Overrides statistic 3.

2 Prints means and standard deviations for each VARIABLES= list with number of valid cases.

3 Prints correlation matrix if one or more correlation coefficients are incalculable.

4 Plots standardized residuals against the sequence of cases in a file; lists observed and predicted values, and residuals.

5 Computes the Durbin-Watson statistics for residuals.

6 Plots standardized residuals against predicted Y values.

7 Prints correlation matrix and number of cases on which it is based.

Limitations (p. 356; p. 223†)

- A maximum of 10 VARIABLES= lists is allowed.
- A maximum of 50 REGRESSION design statements is allowed.
- A maximum of 100 variables is allowed per VARIABLES= list. A maximum of 200 variable names is allowed on the combined VARIABLES= lists.
- A maximum of 100 different variables is allowed on a single REGRESSION design statement.

```
1
REPORT          16
FORMAT= HDSPACE {(3)|(in)} CHDSPACE {(11)|(in)} BRKSPACE {(2)|(in)}

FTSPACE {(1)|(in)} LENGTH {(0,0)|(t,b)} MARGINS {(1,132)|(l, r )}

{NOLIST|LIST} {NOTOTAL|TOTAL} /

AGGREGATE= {TRUE|COMP} {NODUMP|DUMP {(10)|(l n)}} /

STRING= stringname (varname (width) 'literal' . . .

varname (width) 'literal')  stringname . . . /

VARS= varname {(VALUE)|(LABEL)|(DUMMY)} 'column head' (width)  . . varname . . . /

MISSING = {VAR|NONE|LIST (varlist {1|n} )} /

{HEAD=|or|LHEAD=|CHEAD=|RHEAD=} 'line 1' 'line 2'  . . /

{FOOT=|or|LFOOT=|CFOOT=|RFOOT=} 'line 1' 'line 2'  . . /

BREAK= varlist 'breaktitle' (width) {(VALUE)|(LABEL)}

{(NOTITLE)|(TITLE)} {ISKIP {(2)|(ln)} |(PAGE)} /

SUMMARY= RMISS {(–999)|(n)} agg . . agg title'

(break column number) (varname (decimal digits)) .

agg CONTINUE (varname (decimal digits))  . . /
```

All specifications except those shown below are optional (the keyword DEFAULT is not used with any other specification):

```
1
REPORT          16
{FORMAT= DEFAULT/|AGGREGATE= DEFAULT/}

VARIABLES= varlist/

BREAK= varlist/

SUMMARY= agg .  agg/
```

Reference (p. 1†)

Aggregate Statistics (p. 15†)
The following statistics (agg) may be specified with SUMMARY=:

VALIDN*	MEAN*	MAX*
VARIANCE*	STDEV*	SKEWNESS*
SUM*	MIN*	KURTOSIS*

PCTGT(n) PCTLT(n) percent greater/less than n
PCTBTN (n1, n2) percent between n1 and n2
ABFREQ (min, max) RELFREQ (min, max) frequency counts/ percentages within range
MEDIAN (min, max) MODE (min, max) median/modal value within range
DIVIDE (agg (var) agg (var) [factor]) divide first agg by second [and multiply by factor]
PCT (agg (var) agg (var) percent of first over second
SUBTRACT (agg (var) agg (var)) subtract second from first
ADD (agg (var) . . . agg (var)) GREAT (agg (var) . . . agg (var))
LEAST (agg (var) . . . agg (var)) AVERAGE (agg (var) . . . agg (var))

*These statistics may be used in the composite functions, as may numeric constants.

Column widths
Default column width is 9 for an alphanumeric or numeric variable, 20 if LABEL is specified, or the length of a string.

Special arguments
)PAGE prints the page number and)DATE prints mm/dd/yy in heads or running feet.

Options None

Statistics None

Limitations (p. 25†)

- A maximum of 500 variables may be named in the VARS= command.
- A maximum of 10 dummy variables may be defined.
- A maximum of 50 strings may be defined.
- The maximum width under FORMAT= is 132 characters.
- String variables are invalid with AGGREGATE=.
-)PAGE can occur only once in either the head or foot.
- There is a maximum of 20 MODE, MEDIAN, ABFREQ and RELFREQ requests (combined) per summary group.
- There is a maximum of 20 PCTGT, PCTLT, and PCTBTN requests (combined) per summary group.
- If both FORMAT= and AGGREGATE= are specified, the limitations of each apply to the other.

Note: AGGREGATE= is not available in RELEASE 8.0.

```
1
SCATTERGRAM     16
varlist1 {(min,max)|(LOWEST,HIGHEST)} varlist2
[WITH] varlist3  ./varlist . . .
```

Reference (p. 293; p. 223†)

Options (p. 296; p. 223†)

1 Includes missing data.
2 Excludes missing data listwise.
3 Suppresses printing of variable labels.
4 Suppresses printing of plot grid lines.
5 Prints diagonal grids.
6 Applies a two-tailed test of statistical significance.
7 Scales variable values to integers producing SCATTERGRAM plots on integer-labeled scales.
8 If there is not enough core storage space to process all cases, a plot will be produced containing as many randomly selected cases as possible.

Statistics (p. 297)

1 Pearson's r.
2 r-squared
3 Significance of r
4 Standard error of the estimate
5 Intercept with the vertical axis
6 Slope

Limitations (p. 297)

- A maximum of 100 variables may be named or implied.
- A maximum of 25 VARIABLES= lists may be entered.

```
Independent Samples
1
T-TEST          16
GROUPS= {varname {(1,2)|(value)|(value1 value2)} |n1 n2} VARIABLES=varlist

Paired Samples
1
T-TEST          16
PAIRS=varlist [WITH varlist] /varlist
```

Reference (p. 271; p. 224†)

Options (p. 273)

1 Includes missing data.
2 Excludes missing data listwise.
3 Suppresses printing of variable labels.

Statistics No optional statistics.

Limitations (p. 273; p. 224†)

- Only one GROUPS=, one VARIABLES=, and one PAIRS= specifications may be entered.